

Blind background prediction using a bifurcated analysis scheme

J. Nix, J. Ma, G.N. Perdue, Y. Wah

Enrico Fermi Institute, University of Chicago, Chicago, Illinois 60637, USA

(Dated: February 1, 2007)

A technique for background prediction using data, but maintaining a closed signal box is described. The result is extended to two background sources. Conditions on the applicability under correlated cuts are described.

PACS numbers:

In this paper we describe a bifurcation analysis procedure for data driven background prediction under the conditions of a blinded signal box. The procedure uses the application of inverse cuts to properly measure the veto power of the different sets of cuts while not opening the signal box. This technique was first developed for a single background source in the stopped K^+ experiments E787 and E949 at Brookhaven [?]. The work in this paper was inspired by the use of the bifurcation technique in the E391 experiment [?].

We begin with a derivation of the bifurcation analysis in the case of one background source and uncorrelated cuts. We then extend this to two background sources and a simple model of correlation between cuts. We cover the various derivations with a fair amount of algebraic detail. Throughout this paper the method will be applied to a toy model of a background prediction. We utilize the Mathematica software package to simulate this system [?].

ONE BACKGROUND CASE

We begin discussing this method in the case of a single background source. Here a collection of setup cuts have been applied which eliminate all other sources of background. We then want to know the amount of background in the signal region when we apply the cuts A and B, which we refer to as the bifurcation cuts. The number of events we observe will be determined by the number of events before applying the cuts A and B (after applying the setup cuts) and the cut survival probability (CSP).

$$N_{\text{bkg}} = N_0 P(AB) \quad (1)$$

If we consider events to lie in a multi-dimensional space with a dimension corresponding to every variable on which we can cut. Our set of cuts defines a multidimensional signal box which we wish to keep blind. If two cuts show no correlation in the events that they cut, this implies that these two cuts are orthogonal in this space. A diagram of this situation is shown in Figure ?? . The CSP can then be decomposed into $P(AB) = P(A)P(B)$.

$$N_{\text{bkg}} = N_0 P(A)P(B) \quad (2)$$

This can be expanded into

$$N_{\text{bkg}} = \frac{N_0^2 P(A)P(B)P(\bar{A})P(\bar{B})}{N_0 P(\bar{A})P(\bar{B})}. \quad (3)$$

Then we can calculate this from data based on the number of observed events in the signal box under the different cut conditions.

$$N_{A\bar{B}} = N_0 P(A)P(\bar{B}) \quad (4)$$

$$N_{\bar{A}B} = N_0 P(B)P(\bar{A}) \quad (5)$$

$$N_{\bar{A}\bar{B}} = N_0 P(\bar{A})P(\bar{B}) \quad (6)$$

Where $N_{A\bar{B}}$ is the number of background events observed with the application of cut A and the inverse of cut B. $N_{\bar{A}B}$ is the observed background events with the inverse of cut A and cut B applied. $N_{\bar{A}\bar{B}}$ is the count when the inverse of both A and B are applied. All of these values are outside the signal box defined in the multi-dimensional cut space allowing us to predict the background without opening the box.

$$N_{\text{bkg}} = \frac{N_{A\bar{B}}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} \quad (7)$$

The procedure goes as follows. First, apply setup cuts to the data, the number of events in the signal box is N_0 . The setup cuts should remove any background sources other than the ones that we are studying. We then apply the group of cuts A while requiring that events don't pass the set of cuts B. By applying the inverse of B, we are looking at events which are outside the signal box in the multidimensional space. We count the number of events which pass these sets of cuts, $N_{A\bar{B}}$. We then do the same procedure in reverse applying the set of cuts B and the inverse of A to find $N_{\bar{A}B}$. Finally, we apply the inverse of both cuts A and B to find $N_{\bar{A}\bar{B}}$. These values are combined to produce the background prediction.

THE MODEL

Each event consists of four variables. Two kinematic variables, p and x which are used to describe the signal region and two cut variables, a and b which will be used to define the cuts. All of these variables range from 0 to 1.

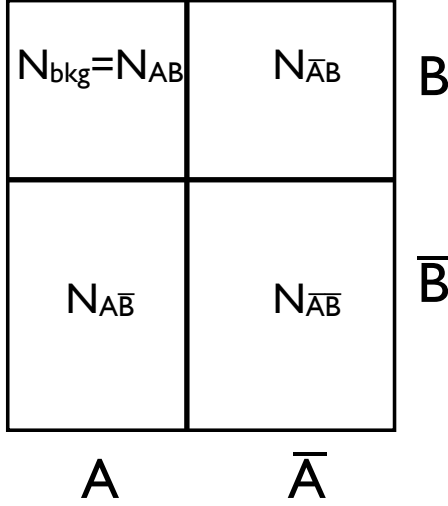


FIG. 1: Schematic of background distribution in the cut space.

We define 2 different types of events: background 1 and background 2. The distribution of each variable for the two background types is shown in Table ??.

Background	x	p
1	$y = 1 \ x \in [0, 1]$ $y = 0 \ x \notin [0, 1]$	$y = 1/x \ p \in [0, x]$ $y = 0 \ p \notin [0, x]$
2	$y = 1 \ x \in [0, 1]$ $y = 0 \ x \notin [0, 1]$	$y = 1/(1-x) \ p \in [0, 1-x]$ $y = 0 \ p \notin [0, 1-x]$

Background	a	b
1	$y = 1 \ a \in [0, 1]$ $y = 0 \ a \notin [0, 1]$	$y = 1 - b \ b \in [0, 1]$ $y = 0 \ b \notin [0, 1]$
2	$y = 1 - a \ a \in [0, 1]$ $y = 0 \ a \notin [0, 1]$	$y = 1 \ b \in [0, 1]$ $y = 0 \ b \notin [0, 1]$

TABLE I: Default variables for each event type in the Toy Model.

We define our cuts on variables a and b as

$$A = (a > 0.5) \quad (8)$$

$$B = (b > 0.5) \quad (9)$$

A and B are true or false statements. If they are false the event is cut. With the cut points defined, we can then calculate the cut survival probability (CSP) for each event type which we represent as $P(A)$ or $P(B)$. In this toy model the CSP's can be calculated analytically because we know the underlying distributions. These values are shown in Table ??.

One Background in the Toy Model

For this section we discuss the case of a single significant background. Our background prediction is given

Event Type	$P(A)$	$P(B)$
Background 1	0.5	0.25
Background 2	0.25	0.5

TABLE II: Cut survival probabilities for each event type in the Toy Model.

N_0	2236 ± 47
$N_{A\bar{B}}$	831 ± 29
$N_{\bar{A}B}$	280 ± 17
$N_{\bar{A}\bar{B}}$	869 ± 29
Predicted Background	267.8 ± 20.6
Observed Background	256 ± 16

TABLE III: Single background study for the Toy Model with only Background 1.

by Eqn ??. We generated 1×10^4 Background 1 events over the whole range of kinematic variables. This leaves us with ≈ 2200 background events in the signal region before applying cuts A and B. In Table ??, we show the observed number of events for each combination of cuts, the predicted background, and the observed background after applying both cut A and B. The predicted background of 267.8 ± 20.6 agrees well with the 256 ± 16 observed background events.

TWO BACKGROUND CASE

Derivation

The previous derivation applied in the case of a single background source. However, if the background is made up of two different background sources, $N_0 = N_1 + N_2$, which have different cut survival probabilities then it is not correct. If that is the case then

$$N_{\text{bkg}} = N_1 P_1(A) P_1(B) + N_2 P_2(A) P_2(B) \quad (10)$$

$$N_{A\bar{B}} = N_1 P_1(A) P_1(\bar{B}) + N_2 P_2(A) P_2(\bar{B}) \quad (11)$$

$$N_{\bar{A}B} = N_1 P_1(B) P_1(\bar{A}) + N_2 P_2(B) P_2(\bar{A}) \quad (12)$$

$$N_{\bar{A}\bar{B}} = N_1 P_1(\bar{A}) P_1(\bar{B}) + N_2 P_2(\bar{A}) P_2(\bar{B}) \quad (13)$$

Then our previous calculation of the background has a cross term introduced. We wish to find the correction to the one background solution. We begin by substituting the above definitions into the solution for the one background case, Eqn ??.

$$\frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} = \frac{1}{N_{\bar{A}\bar{B}}} [N_1 P_1(A) P_1(\bar{B}) + N_2 P_2(A) P_2(\bar{B})] \times [N_1 P_1(B) P_1(\bar{A}) + N_2 P_2(B) P_2(\bar{A})] \quad (14)$$

We can expand the numerator into

$$\begin{aligned}
N_{AB}N_{\bar{A}\bar{B}} &= N_1^2 P_1(A)P_1(\bar{A})P_1(B)P_1(\bar{B}) \\
&\quad + N_1 N_2 [P_1(A)P_2(\bar{A})P_2(B)P_1(\bar{B}) \\
&\quad + P_2(A)P_1(\bar{A})P_1(B)P_2(\bar{B})] \\
&\quad + N_2^2 P_2(A)P_2(\bar{A})P_2(B)P_2(\bar{B}).
\end{aligned} \tag{15}$$

We multiply N_{bkg} by $N_{\bar{A}\bar{B}}$ to allow us to find the difference of Eqn. ?? and N_{bkg} .

$$\begin{aligned}
N_{\text{bkg}}N_{\bar{A}\bar{B}} &= [N_1 P_1(A)P_1(B) + N_2 P_2(A)P_2(B)] \times \\
&\quad [N_1 P_1(\bar{A})P_1(\bar{B}) + N_2 P_2(\bar{A})P_2(\bar{B})] \\
&= N_1^2 P_1(A)P_1(\bar{A})P_1(B)P_1(\bar{B}) \\
&\quad + N_1 N_2 [P_1(A)P_2(\bar{A})P_1(B)P_1(\bar{B}) \\
&\quad + P_2(A)P_1(\bar{A})P_2(B)P_1(\bar{B})] \\
&\quad + N_2^2 P_2(A)P_2(\bar{A})P_2(B)P_2(\bar{B})
\end{aligned} \tag{16}$$

$$\begin{aligned}
\frac{N_{AB}N_{\bar{A}\bar{B}}}{N_{\bar{A}\bar{B}}} &= N_{\text{bkg}} \\
&\quad + \frac{N_1 N_2}{N_{\bar{A}\bar{B}}} \left[P_1(A)P_2(\bar{A})P_2(B)P_1(\bar{B}) \right. \\
&\quad + P_2(A)P_1(\bar{A})P_1(B)P_2(\bar{B}) \\
&\quad - P_1(A)P_2(\bar{A})P_1(B)P_2(\bar{B}) \\
&\quad \left. - P_2(A)P_1(\bar{A})P_2(B)P_1(\bar{B}) \right]
\end{aligned} \tag{17}$$

The cross term vanishes if one of the following conditions are met

1. $P_1(A)$ or $P_2(A) = 0$ and $P_1(B)$ or $P_2(B) = 0$ In this case each of cuts completely eliminate one of the background sources. However, this condition implies that $N_{\bar{A}\bar{B}}$ is zero which means the correction term is poorly defined.
2. $P_1(A) = P_2(A)$ and $P_1(B) = P_2(B)$ In this case for the purposes of the cuts, the two background sources are the same.
3. $N_1 = 0$ or $N_2 = 0$ Here there is only one background source in the sample.

We can simplify the cross term by rewriting the CSP's of the inverse cuts in terms of the CSP's of the cuts, $P_i(\bar{A}) = 1 - P_i(A)$. Each element of the cross term has

the same structure which can be expanded to

$$\begin{aligned}
P_i(A)P_j(\bar{A})P_k(B)P_\ell(\bar{B}) &= P_i(A)P_k(B)(1 - P_j(A)) \times \\
&\quad (1 - P_\ell(B)) \\
&= P_i(A)P_k(B)(1 - P_j(A) \\
&\quad - P_\ell(B) + P_j(A)P_\ell(B)) \\
&= P_i(A)P_k(B) \\
&\quad - P_i(A)P_j(A)P_k(B) \\
&\quad - P_i(A)P_k(B)P_\ell(B) \\
&\quad + P_i(A)P_j(A)P_k(B)P_\ell(B).
\end{aligned} \tag{18}$$

Summing the elements of the cross term cancels out everything except the terms with two CSP's.

$$\begin{aligned}
\frac{N_{AB}N_{\bar{A}\bar{B}}}{N_{\bar{A}\bar{B}}} &= N_{\text{bkg}} \\
&\quad + \frac{1}{N_{\bar{A}\bar{B}}} (N_1 N_2 (P_1(A)P_2(B) + P_2(A)P_1(B) \\
&\quad - P_1(A)P_1(B) - P_2(A)P_2(B)) \\
&= N_{\text{bkg}} - \frac{N_1 N_2 (P_2(A) - P_1(A))(P_2(B) - P_1(B))}{N_{\bar{A}\bar{B}}}
\end{aligned} \tag{19}$$

We can further simplify the cross term by defining $\Delta_A = P_2(A) - P_1(A)$ and $\Delta_B = P_2(B) - P_1(B)$.

$$N_{\text{bkg}} = \frac{N_{AB}N_{\bar{A}\bar{B}}}{N_{\bar{A}\bar{B}}} + \frac{N_1 N_2}{N_{\bar{A}\bar{B}}} \Delta_A \Delta_B \tag{20}$$

It is important to note that the correction term is not the contribution of a particular source to the background prediction.

Properties of the Two Background Solution

This solution has the reasonable property that it is symmetric with respect to the definitions of the cuts A and B and the backgrounds 1 and 2.

The correction term can be either positive or negative. However, the total N_{bkg} will not be negative. The correction term will have it's maximum negative value when $\Delta_A = 1$ and $\Delta_B = -1$ or $\Delta_A = -1$ and $\Delta_B = 1$. Under these conditions, $N_{\bar{A}\bar{B}} = 0$ and N_{bkg} is undefined. We therefore want to study N_{bkg} 's behavior as we approach this limit. We begin by setting $\Delta_B = -1$ and studying the limit as $\Delta_A \rightarrow 1$.

The condition that $\Delta_B = -1$ sets what values the CSP's of the B can take.

$$P_1(B) = P_2(\bar{B}) = 0 \tag{21}$$

$$P_2(B) = P_1(\bar{B}) = 1 \tag{22}$$

Substituting these values into Eqns. ??-??, we find

$$N_{AB} = N_2 P_2(A), \quad (23)$$

$$N_{\bar{A}B} = N_1 P_1(\bar{A}) = N_1(1 - P_1(A)), \quad (24)$$

$$N_{A\bar{B}} = N_2 P_2(\bar{B}) = N_2(1 - P_2(A)). \quad (25)$$

We then substitute these values into Equation ?? and sum the two terms

$$\begin{aligned} N_{\text{bkg}} &= \frac{N_1(P_2(A) - P_2(A)P_1(A) - P_2(A) + P_1(A))}{1 - P_2(A)} \\ &= \frac{N_1 P_1(A)(1 - P_2(A))}{1 - P_2(A)} = N_1 P_1(A). \end{aligned} \quad (26)$$

As $\Delta_A \rightarrow 1$, $P_1(A) \rightarrow 0$. Therefore the N_{bkg} goes to 0. This indicates that N_{bkg} never has a non-physical negative value.

Interpretation of Two Background Solution

It is counterintuitive that two backgrounds cannot be combined simply. This can best be understood as the second background introducing an implicit correlation.

Let's consider the extreme case of dividing the original background in two, one background where $P(A) = 1$ and the other with $P(A) = 0$. This means $\Delta_A = \pm 1$. Whether this changes the result depends on the value of Δ_B . If $\Delta_B \neq 0$ then there is a correction to the original prediction. The fact that $\Delta_B \neq 0$ implies there is already a correlation between cut A and cut B.

If we instead consider two backgrounds which each individually have no correlation between cut A and cut B, but do have different cut survival probabilities combining them introduces a correlation. If $P_1(A) = 0.75$ and $P_1(B) = 0.75$, while $P_2(A) = 0.25$ and $P_2(B) = 0.25$ then the resulting combination of the two backgrounds will have a correlation such that events which survive cut A are likely to survive cut B, while events don't survive cut A are likely to not survive cut B. Therefore there is a correlation, even though the individual backgrounds are uncorrelated.

It is important to note that values of N_1 , N_2 , Δ_A , and Δ_B are not directly accessible in data without opening the signal box. There are two options either derive these values from Monte Carlo or from other regions in signal space. N_1 and N_2 generally will require some alternative way of predicting one of the backgrounds and the value of N_0 , the total number of background after setup cuts. This then raises the question, does determining N_0 bias the analysis. From N_0 and the other observed background numbers, N_{AB} , $N_{\bar{A}B}$, and $N_{A\bar{B}}$, it is possible to

effectively open the box and count N_{bkg} . Determining Δ_A and Δ_B also requires additional input. Their values can be derived from either Monte Carlo or data outside the signal region.

Two Background Toy Model

In our toy model, we can study the effect of multiple background sources by varying the relative strength of a second background. We begin by calculating with the false assumption that there is a single background mode. We vary the relative admixture of Background 2. The total number of events, $N_1 + N_2$, was held constant at 2×10^4 . The discrepancy between the prediction and the observed background increases as the the number of background events from the second source increases.

Secondary Background Correction

We now apply the correction term to the background prediction (Eqn ??). In the case of this toy model, we know the values of N_1 and N_2 because we have set them. In a real analysis, it would be necessary to determine these values through either Monte Carlo studies or studies of different signal regions which are then extrapolated into the signal box. The differences in the cut probabilities, Δ_A and Δ_B , also need to be determined from outside sources. In this model $\Delta_A = -25\%$ and $\Delta_B = 25\%$. Since the probability differences are of opposite signs the correction factor is negative and reduces the predicted background.

In Figure ??, we show the results of keeping the total number of background events the same while increasing the fraction of Background 2 events. Here only the predicted background without correction increases, while the observed and corrected backgrounds remain relatively flat.

CUT CORRELATION

In the derivation of both the one and two background cases, we have assumed that the cuts A and B are uncorrelated. Of course, in real applications, it is unlikely to find two cuts which are perfectly uncorrelated. We would therefore like to find some general figure of merit to determine how the correlation of the introduces errors into the background prediction.

Impact of Cut Correlation

There are a variety of general measures of correlation between variables, such as the statistical correlation, but

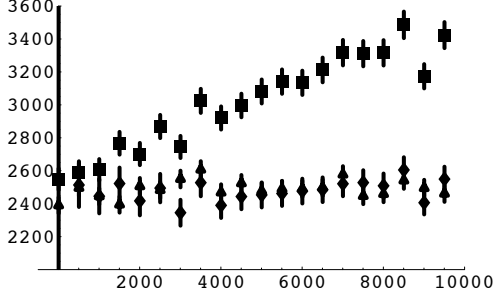


FIG. 2: Predicted and observed background for different admixtures of a Background 2. Squares are predicted background (without second background correction), triangles are the observed background in data, diamonds are the corrected prediction. The golden triangles are the value of the two background correction. The x-axis is the number of generated Background 2 events, the total number of events, $N_1 + N_2$, was held constant at 2×10^4 .

without a model of how the cuts are correlated it is difficult to derive a correction to the background prediction. We describe a case where the cuts have a weak linear correlation to establish the impact on the background prediction. In this model the CSP's of each cut have a small difference in value when the other cut applies. Since the cuts are correlated, we must always specify the condition of the other cut when quoting a cut's survival probability.

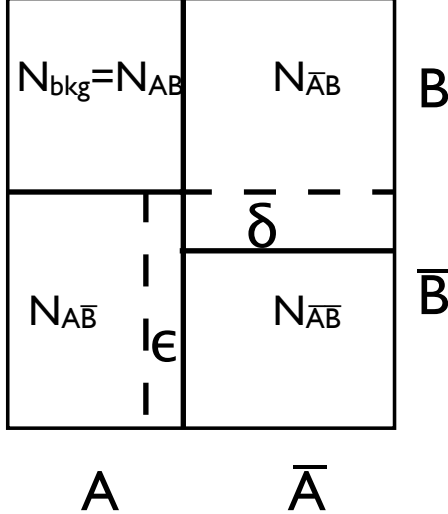


FIG. 3: Cut space with correlated cuts.

We begin by specifying the background values in terms of the CSP's, which now specify the dependence on both cut conditions.

$$N_{\text{bkg}} = N_0 P(A|B) P(B) = N_0 P(A) P(B|A) \quad (27)$$

$$N_{A\bar{B}} = N_0 P(A|\bar{B}) P(\bar{B}) = N_0 P(A) P(\bar{B}|A) \quad (28)$$

$$N_{\bar{A}B} = N_0 P(\bar{A}|B) P(B) = N_0 P(\bar{A}) P(B|\bar{A}) \quad (29)$$

$$N_{\bar{A}\bar{B}} = N_0 P(\bar{A}|\bar{B}) P(\bar{B}) = N_0 P(\bar{A}) P(\bar{B}|\bar{A}) \quad (30)$$

We proceed in the same fashion as for the two background case and substitute these definitions into the solution (Equation 7) for the single background uncorrelated case.

$$\frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} = \frac{N_0^2 P(A|\bar{B}) P(\bar{B}) P(\bar{A}|B) P(B)}{N_0 P(\bar{A}|\bar{B}) P(\bar{B})} \quad (31)$$

We are interested in the case where the correlations are small, so we define

$$P(A|\bar{B}) = P(A|B) - \epsilon \quad (32)$$

$$P(\bar{A}|\bar{B}) = P(\bar{A}|B) + \epsilon \quad (33)$$

$$P(B|\bar{A}) = P(B|A) - \delta \quad (34)$$

$$P(\bar{B}|\bar{A}) = P(\bar{B}|A) + \delta. \quad (35)$$

The corrections ϵ and δ should be small. What we mean by small will be defined at the end of the derivation by what values are necessary for the corrections which are first order in ϵ and δ to be negligible. We substitute these definitions into Equation 31.

$$\frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} = \frac{N_0 P(\bar{A}|B) (P(A|B) - \epsilon) P(B)}{P(\bar{A}|B) + \epsilon} \quad (36)$$

$$\frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} = \frac{N_0 (P(A|B) P(B) - \epsilon P(B))}{1 + \frac{\epsilon}{P(\bar{A}|B)}} \quad (37)$$

If we assume the ϵ term in the denominator are small we expand this result as

$$\begin{aligned} \frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} &\approx (N_0 (P(A|B) P(B) - \epsilon P(B))) \\ &\times (1 - \frac{\epsilon}{P(\bar{A}|B)} + \frac{\epsilon^2}{P(\bar{A}|B)^2} + \mathcal{O}(\epsilon^3)). \end{aligned} \quad (38)$$

Multiplying this out and keeping the second order terms of ϵ gives equation

$$\begin{aligned} \frac{N_{A\bar{B}} N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} &= N_0 P(A|B) P(B) - \epsilon N_0 (P(B) + \frac{P(A|B) P(B)}{P(\bar{A}|B)}) \\ &+ \epsilon^2 N_0 (\frac{P(B)}{P(\bar{A}|B)} + \frac{P(A|B) P(B)}{P(\bar{A}|B)^2}) \end{aligned} \quad (39)$$

The first term with no ϵ factors is N_{bkg} . The condition for the correlations to have a negligible impact on

our background prediction is that the ϵ terms be much smaller than $\frac{N_{AB}N_{\bar{A}\bar{B}}}{N_{\bar{A}\bar{B}}}$.

$$N_{\text{bkg}} = \frac{N_{A\bar{B}}N_{\bar{A}B}}{N_{\bar{A}\bar{B}}} + \epsilon N_0 P(B) \left(1 + \frac{P(A|B)}{P(\bar{A}|B)}\right) - \epsilon^2 N_0 \frac{P(B)}{P(\bar{A}|B)} \left(1 + \frac{P(A|B)}{P(\bar{A}|B)}\right) \quad (40)$$

These terms however, require opening the signal box to know the correct values of the CSP's. We can however approximate these values with less knowledge, under the assumption that the number of events in the signal box is small.

$$P(B) = \frac{N_{AB} + N_{\bar{A}B}}{N_0} \quad (41)$$

$$P(B) \approx \frac{N_{\bar{A}B}}{N_0} \quad (42)$$

$$\frac{P(A|B)}{P(\bar{A}|B)} = \frac{\frac{N_{AB}}{N_B}}{\frac{N_{\bar{A}B}}{N_B}} \approx \frac{N_{\text{pred}}}{N_{\bar{A}B}} \quad (43)$$

These approximations give us first order correction of

$$\text{1st Order} = \epsilon N_{\bar{A}B} \left(1 + \frac{N_{\text{pred}}}{N_{\bar{A}B}}\right) \quad (44)$$

We introduce a correlation between the a and b variables in Background 1. We add a term linearly dependent on b to a , and then rescale a to keep it between 0 and 1 and to reduce the change in background due to just the change in the average value of a .

$$a = \frac{1 - \sqrt{t} + \epsilon' b}{1 + \epsilon'}, t \in [0, 1], b \in [0, 1] \quad (45)$$

The variable ϵ' is the knob we use to tune the correlation. It is closely related to the variable ϵ that is defined in Equations ?? and ?? as is shown in Figure ?. For easier comparison to that technote, we describe the variation of the model in terms of the ϵ variable.

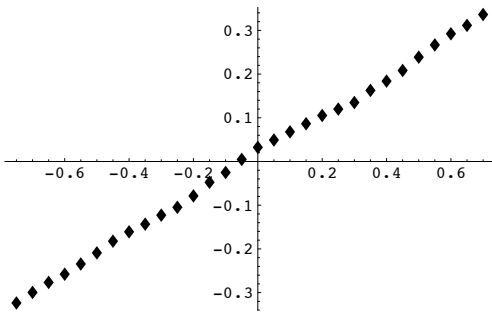


FIG. 4: The value of ϵ (Eqn. ??) as a function of ϵ' (Eqn. ??).

We show the predicted and observed background in Figure ?. As ϵ' increases the background in data increases, as the correlation increases the average value a , while the predicted background decreases.

$$C_\epsilon = N_0 \left(\epsilon(P(B) + \frac{P(A|B)P(B)}{P(\bar{A}|B)}) - \epsilon^2 \left(\frac{P(B)}{P(\bar{A}|B)} + \frac{P(A|B)P(B)}{P(\bar{A}|B)^2} \right) \right). \quad (46)$$

In Figure ??, we show the prediction with C_ϵ added. It improves the agreement for a fairly wide range of ϵ .

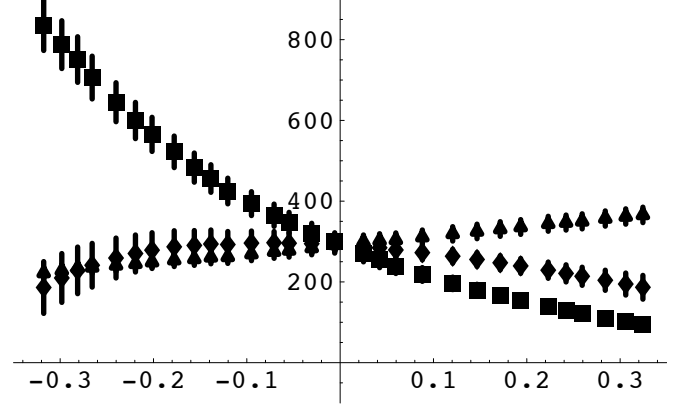


FIG. 5: Predicted and observed background for different levels of correlation between a and b . Triangles are data background, squares are predicted, and diamonds are predicted plus second order ϵ correction (Eqn. ??). The x-axis is ϵ .

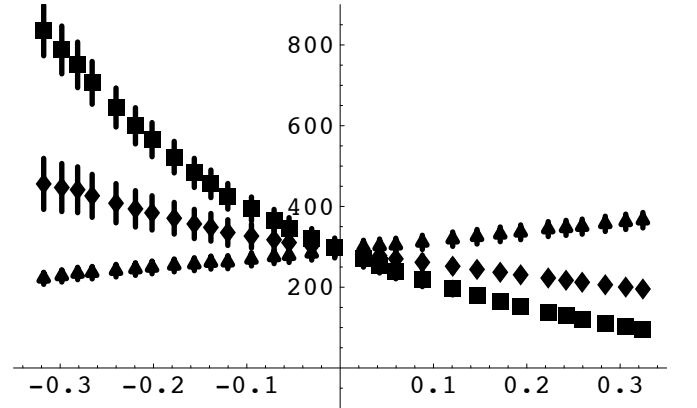


FIG. 6: Predicted and observed background for different levels of correlation between a and b . Triangles are data background, squares are predicted, and diamonds are predicted plus practical ϵ correction (Eqn. ??). The x-axis is ϵ .

DISCUSSION

The bifurcation analysis technique allows us to produce data driven background predictions while still maintaining a blind analysis. In this paper we have shown how to

extend the bifurcation analysis to the case of two background sources and correlated cuts. The primary purpose of both of these is to estimate errors on the one background, uncorrelated cut analysis.

The corrections for a second background source is effective for any level of secondary background. It does require information beyond which is available directly from a blind analysis. It requires a combination of Monte Carlo information about the relative strengths of the two backgrounds and how the cut survival probabilities vary between the two backgrounds. It also requires knowledge of N_0 , the number of events in the signal box after the setup cuts.

Correlations between the two cuts cannot be handled as easily. Even when the correlation between cuts is linear, the effects on the background prediction are non-linear. Therefore, special care must be taken when selecting the cuts for the bifurcation analysis to avoid correlation.

One particular aspect of the impact of the prediction on the cut correlation is its dependence on the value of N_0 . This leads to two competing forces in optimizing the division of cuts into the setup cuts and the bifurcation cuts. From the perspective of minimizing statistical error in the bifurcation prediction, one would like a large value for N_0 with powerful cuts for cut A and B , so that the statistical errors on the terms of Eqn. ?? are small. On the other hand, a large N_0 means the prediction is sensitive to small correlations between the cuts.

-
- S. Adler et al. "Evidence for the Decay $K^+ \rightarrow \pi^+ \nu \bar{\nu}$ ", Phys.Rev.Lett. 79 (1997) 2204
 - J. Nix et al. "First Search for $K_L \rightarrow \pi^0 \pi^0 \nu \bar{\nu}$ ",
 - Wolfram Research, Inc., Mathematica, Version 5.2, Champaign, IL (2005).